Vorlesung 7a

Der Zentrale Grenzwertsatz

Teil 2

Ein Beweis

(Buch S. 78-80)

Seien $X_1, X_2, \ldots$ unabhängig und identisch verteilt

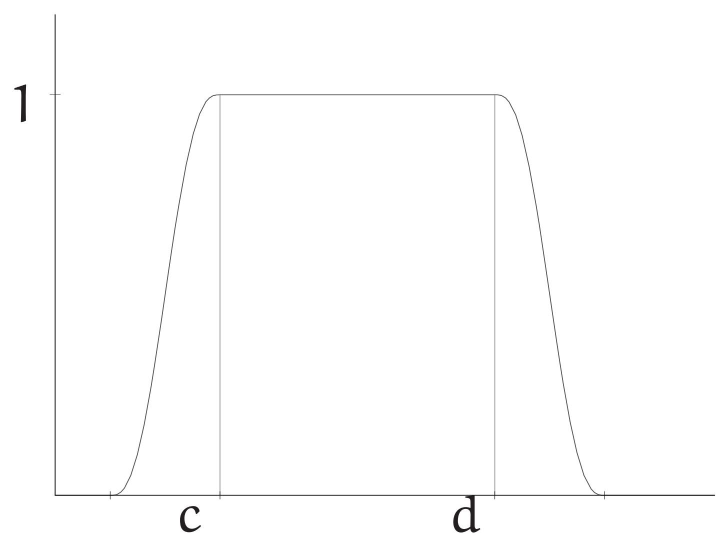
mit Erwartungswert 0 und Varianz 1.

Die Behauptung ist :

$$\mathbf{E}\Big[\mathbf{1}_{[c,d]}\Big(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\Big)\Big] \underset{n\to\infty}{\longrightarrow} \mathbf{E}[\mathbf{1}_{[c,d]}(Z)]$$

mit standard-normalverteiltem Z.

Weil man Indikatorfunktionen $\mathbf{1}_{[c,d]}$
durch “glatte” Funktionen h approximieren kann, reicht es
(wie man zeigen kann, mehr dazu später),
diese Konvergenz nur für glatte h zu beweisen.



Lemma

(Buch S. 78)

Sei $h : \mathbb{R} \to \mathbb{R}$ dreimal stetig differenzierbar und seien h', h'' und h''' beschränkt. Dann gilt

$$\mathbf{E}\left[h\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right)\right] \xrightarrow[n\to\infty]{} \mathbf{E}[h(Z)]$$

Zum Beweis des Lemmas: *

Die Hauptidee besteht darin, eine Folge von unabhängigen standard-normalverteilten Zufallsvariablen $(Z_1, Z_2, \ldots)$ ins Spiel zu bringen, die zusammen mit $(X_1, X_2, \ldots)$ ein zufälliges Paar von Folgen bilden.

Dabei seien alle $Z_1, Z_2, \ldots, X_1, X_2, \ldots$ unabhängig.

Wir wissen schon, dass gilt:

$$\mathbf{E}[h(Z)] = \mathbf{E}\left[h\left(\frac{Z_1 + \cdots + Z_n}{\sqrt{n}}\right)\right].$$

*nach einer Idee von J. Lindeberg, * 1876, † 1932

Außerdem ergibt sich

mit der Linearität des Erwartungswertes:

$$\mathbf{E}\left[h\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right]-\mathbf{E}\left[h\left(\frac{Z_1+\cdots+Z_n}{\sqrt{n}}\right)\right]$$

$$=\mathbf{E}\left[h\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)-h\left(\frac{Z_1+\cdots+Z_n}{\sqrt{n}}\right)\right].$$

Es reicht also zu zeigen, dass letzteres

für $n \to \infty$ gegen Null konvergiert.

Eine clevere Idee ist es jetzt, die Differenz als Teleskopsumme darzustellen:

$$h\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right) - h\left(\frac{Z_1 + \cdots + Z_n}{\sqrt{n}}\right)$$

$$= \sum_{i=1}^{n} \left(h\left(\frac{X_1 + \cdots + X_{i-1} + \color{red}{X_i} + Z_{i+1} + \cdots + Z_n}{\sqrt{n}}\right)\right.$$

$$\left. - h\left(\frac{X_1 + \cdots + X_{i-1} + \color{red}{Z_i} + Z_{i+1} + \cdots + Z_n}{\sqrt{n}}\right)\right)$$

Taylorentwicklung ergibt:

$$= \sum_{i=1}^{n} \frac{X_i - Z_i}{\sqrt{n}} h' \left(\frac{X_1 + \cdots + X_{i-1} + Z_{i+1} + \cdots + Z_n}{\sqrt{n}} \right)$$

$$+ \sum_{i=1}^{n} \frac{X_i^2 - Z_i^2}{2n} h'' \left(\frac{X_1 + \cdots + X_{i-1} + Z_{i+1} + \cdots + Z_n}{\sqrt{n}} \right)$$

$$+ \sum_{i=1}^{n} \left(\frac{X_i^3}{6n^{3/2}} h'''(Y_i) - \frac{Z_i^3}{6n^{3/2}} h'''(\tilde{Y}_i) \right)$$

mit passenden Zwischenstellen Y_i, $\tilde{Y}_i$.

Wir nehmen hier der Einfachheit halber an:

$$\mathbf{E}[|X_1|^3] < \infty.$$

(Der Fall ohne diese Zusatzbedingung

ist im Buch S. 78 abgehandelt.)

Ist C eine obere Schranke von $|h'''|$, so folgt

$$\begin{aligned} &\Big|\mathbf{E}\Big[h\Big(\frac{X_1+\cdots+X_n}{\sqrt{n}}\Big)-h\Big(\frac{Z_1+\cdots+Z_n}{\sqrt{n}}\Big)\Big]\Big| \\ &\leq n\frac{1}{6n^{3/2}}(\mathbf{E}[|X_1|^3]+\mathbf{E}[|Z_1|^3])C \longrightarrow 0, \end{aligned}$$

denn die Erw. werte der ersten beiden Summen sind Null

wegen $\mathbf{E}[X_1]=\mathbf{E}[Z_1]=0$ und $\mathbf{E}[X_1^2]=\mathbf{E}[Z_1^2]=1$,

zusammen mit der Unabhängigkeit der X_i, Z_i

und der Produktformel

für die Erwartungswerte unabhängiger Zufallsvariabler. $\square$

Wir folgern jetzt die Ausage des Zentralen Grenzwertsatzes

aus dem Lemma,

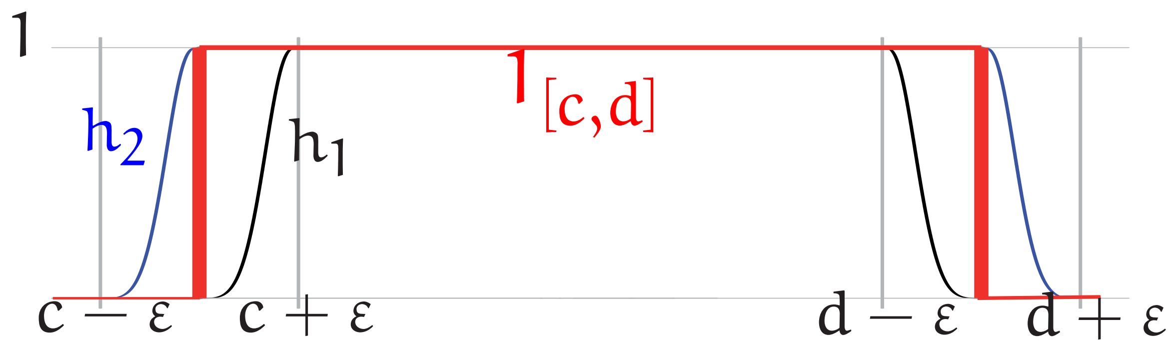
uind zwar mittels der vesrprochenen Approximation

der Indikatorfunktionen $\mathbf{1}_{[c,d]}$

durch hinreichend glatte Funktionen h.

Dabei verwenden wir auch den

Satz über die Mononotonie des Erwartungswertes.



$$\mathbf{1}_{[c+\varepsilon,d-\varepsilon]} \le h_1 \le \mathbf{1}_{[c,d]} \le h_2 \le \mathbf{1}_{[c-\varepsilon,d+\varepsilon]}$$

$$\begin{gathered}
\mathbf{P}(c+\varepsilon \le Z \le d-\varepsilon) \\
\le \mathbf{E}[h_1(Z)] = \lim_{n\to\infty} \mathbf{E}\left[h_1\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right] \\
\le \liminf_{n\to\infty} \mathbf{P}\left(c \le \frac{X_1+\cdots+X_n}{\sqrt{n}} \le d\right) \\
\le \limsup_{n\to\infty} \mathbf{P}\left(c \le \frac{X_1+\cdots+X_n}{\sqrt{n}} \le d\right) \\
\le \lim_{n\to\infty} \mathbf{E}\left[h_2\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right] = \mathbf{E}[h_2(Z)] \\
\le \mathbf{P}(c-\varepsilon \le Z \le d+\varepsilon)
\end{gathered}$$

$$\mathbf{P}(c + \varepsilon \leq Z \leq d - \varepsilon)$$

$$\leq \liminf_{n\to\infty} \mathbf{P}\left(c \leq \frac{X_1+\cdots+X_n}{\sqrt{n}} \leq d\right)$$
$$\leq \limsup_{n\to\infty} \mathbf{P}\left(c \leq \frac{X_1+\cdots+X_n}{\sqrt{n}} \leq d\right)$$

$$\leq \mathbf{P}(c - \varepsilon \leq Z \leq d + \varepsilon).$$

Da Z eine Dichte besitzt, gilt

$\mathbf{P}(c \pm \varepsilon \leq Z \leq d \mp \varepsilon) \to \mathbf{P}(c \leq Z \leq d)$ für $\varepsilon \to 0$,

und die Behauptung folgt. $\square$